DOCUMENT RESUME

ED 384 664 TM 023 958

AUTHOR Mislevy, Robert J.
TITLE Some Formulas for Use with Bayesian Ability Estimates.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-93-3
PUB DATE Jan 93
NOTE 21p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Ability; *Bayesian Statistics; Computer Software; Estimation (Mathematics); *Item Response Theory; *Population Distribution; *Statistical Distributions; *Test Theory
IDENTIFIERS *Ability Estimates

ABSTRACT

Relationships between Bayesian ability estimates and the parameters of a normal population distribution are derived in the context of classical test theory. Analogies are provided for use as approximations in work with item response theory (IRT). The following issues are addressed: (1) the relationship between the distribution of the latent ability variable in a population and the distribution of ability estimates; (2) how to proceed in calculating the Bayesian estimate if the population distribution is not known; and (3) how to estimate the distributions of specified subpopulations when Bayesian ability estimates have been calculated in accordance with a common population distribution. Exact relationships are derived to address these questions in the context of classical test theory, assuming normally distributed abilities and errors. The analogues that are offered are for a not-uncommon IRT context in which the researcher has software to calculate the Bayesian IRT estimates for individuals under the assumption of a normal population distribution, but possesses neither values of the population parameters nor software with which to estimate them. (Contains 1 figure and 11 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
H. I. BRAUN
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

RESEARCH REPORT

SOME FORMULAS FOR USE WITH BAYESIAN ABILITY ESTIMATES

Robert J. Mislevy

ED 384 664

BEST COPY AVAILABLE

Educational Testing Service
Princeton, New Jersey
January 1993

TM023958

Some Formulas for Use with Bayesian Ability Estimates

Robert J. Mislevy

Educational Testing Service

Copyright © 1993. Educational Testing Service. All rights reserved.

Some Formulas for Use with Bayesian Ability Estimates

Abstract

Relationships between Bayesian ability estimates and the parameters of a normal population distribution are derived in the context of classical test theory. Analogues are provided for use as approximations in work with item response theory. The following questions addressed:

- What is the relationship between the distribution of the latent ability variable in a population, and the distribution of ability estimates?
- Because calculating Bayesian estimates typically requires knowing the population distribution, how should one proceed if it is not known?
- What if Bayesian ability estimates have been calculated in accordance with a common population distribution, but it is later desired to estimate the distributions of specified subpopulations?

Key words: Bayesian estimation, classical test theory, item response theory

Some Formulas for Use with Bayesian Ability Estimates

Introduction

From the time of Truman Kelley (1923), Bayesian ability estimates have often been used in educational testing. Reasons for doing so range from Novick's theoretical arguments for Bayesian inference in general (e.g., Novick and Jackson, 1974) to a more practical desire to obtain finite ability estimates for all examinees in the context of item response theory (IRT). This paper provides some formulas for practical work with Bayesian ability estimates, focusing on the following questions:

1. What is the relationship between the distribution of the latent ability variable in a population and the distribution of Bayesian ability estimates?

2. Because calculating Bayesian estimates typically requires knowing the population distribution, how should one proceed if it is not known?

3. What if Bayesian ability estimates have been calculated using a common population distribution, but it is later desired to estimate the distributions of specified subpopulations?

Exact relationships are derived to address these questions in the context of classical test theory, assuming normally distributed abilities and errors. Analogues are offered as computing approximations in a not-uncommon IRT context: A researcher has software to calculate Bayesian IRT estimates for individuals under the assumption of a normal population distribution, but possesses neither values of the population parameters nor software with which to estimate them.

Classical Test Theory

Background and Notation

The symbol θ denotes a real-valued latent proficiency variable, assumed to follow a normal distribution in a population of examinees; that is,

$$\theta|\mu,\sigma^2 \sim \mathrm{N}(\mu,\sigma^2)\,. \qquad (1)$$

Under classical test theory (CTT) one observes the value of the manifest variable x, which is the sum of the latent variable and an independent, real-valued error or disturbance term e:

$$x = \theta + e\,. \tag{2}$$

If normality is assumed for the error terms,

$$e \sim \mathrm{N}(0, \sigma_e^2). \tag{3}$$

Equivalently, the conditional distribution of x given θ can be written as

$$x|\theta \sim \mathrm{N}\left(\theta, \sigma_e^2\right). \tag{4}$$

Together, Equations 1 through 3 imply that

$$\theta|\mu, \sigma^2, \sigma_e^2 \sim \mathrm{N}\left(\mu, \sigma^2 + \sigma_e^2\right). \tag{5}$$

When an individual's x is observed, Equation 4 is interpreted as a likelihood function for the unobserved θ, denoted $\ell(\theta|x)$. Under the assumptions outlined previously,

$$\ell(\theta|x) = \mathrm{N}(x, \sigma_e^2)\,, \tag{6}$$

a normal distribution with mean x and variance σ_e^2. The maximum likelihood estimate (MLE) of an examinee's θ, denoted $\hat{\theta}$, is therefore simply x in this context, and the estimation error variance is σ_e^2.

$\mathrm{N}(\mu, \sigma^2)$ is the prior distribution for an examinee's θ value under CTT. It represents what is known about θ before a test score is observed. Suppose μ, σ^2, and σ_e^2 are known. The posterior distribution for an individual's θ after observing x is obtained by Bayes theorem as $\mathrm{p}(\theta|x, \mu, \sigma^2, \sigma_e^2) \propto \ell(\theta|x)\ \mathrm{p}(\theta|\mu, \sigma^2)$. If normality is assumed for e and θ, then the posterior is also normal:

$$\theta|x, \mu, \sigma^2, \sigma_e^2 \sim \mathrm{N}\left(\overline{\theta}, \overline{\sigma}^2\right),$$

with (posterior) variance

$$\begin{aligned} \overline{\sigma}^2 &= \left(\sigma^{-2} + \sigma_e^{-2}\right)^{-1} \\ &= (1 - \rho)\sigma^2, \end{aligned} \tag{7}$$

where ρ is the reliability coefficient, defined by

$$\rho = \frac{\sigma^2}{\sigma^2 + \sigma_e^2}, \tag{8}$$

and with (posterior) mean

$$\begin{aligned}\overline{\theta} &= \rho x + (1-\rho)\mu \\ &= \frac{\sigma^2}{\sigma^2 + \sigma_e^2} x + \frac{\sigma_e^2}{\sigma^2 + \sigma_e^2}\mu.\end{aligned} \tag{9}$$

(see Box and Taio, 1973, pp. 74-75, for a proof). Equations 7 and 9 are familiar as Kelley's (1923) formulas. $\overline{\theta}$ is the Bayes mean, or expectation a posteriori (EAP), estimate of θ for an examinee with observed response x. Because the posterior is normal, $\overline{\theta}$ is also the Bayes modal estimate for θ, or the mode of its posterior.

Question 1: What is the relationship between the distribution of the latent ability variable in a population, and the distribution of ability estimates?

Because the bottom line in test theory is usually inference about individual examinees, attention has focused on obtaining scores for individuals that are optimal in one sense or another. MLEs are consistent and best asymptotically normal estimates of individuals' θs; Bayesian estimates minimize the average squared difference between estimates and true values. A fundamental paradox of test theory is that *the distribution of these "good" estimates of individuals' θs is <u>not</u> a good estimate of the θ distribution* (Lord, 1969; Mislevy, Beaton, Kaplan, and Sheehan, 1992). In the CTT setting described above, both MLEs and Bayesian estimates follow normal distributions. Their means are equal to the mean of θ in the population, but their variances are *not* equal to the variance of θ:

For MLEs,

$$E\left(\hat{\theta}\right) = E(x) = \mu, \tag{10}$$

but

$$Var\left(\hat{\theta}\right) = Var(x) = \sigma^2 + \sigma_e^2. \tag{11}$$

For Bayesian estimates,

$$\begin{aligned} E(\bar{\theta}) &= E[\rho x + (1-\rho)\mu] \\ &= \rho E(x) + (1-\rho)E(\mu) \\ &= \rho\mu + (1-\rho)\mu \\ &= \mu, \end{aligned} \tag{12}$$

but

$$\begin{aligned} Var(\bar{\theta}) &= Var[\rho x + (1-\rho)\mu] \\ &= \rho^2 Var(x) \\ &= \rho^2(\sigma^2 + \sigma_e^2) \\ &= \rho\sigma^2. \end{aligned} \tag{13}$$

The decomposition of variance implied by Equations 7 and 13 should be noted: the variance of θ can be expressed as the sum of the posterior variance (which is the same for all examinees under CTT) and the variance of the Bayes mean estimates:

$$\begin{aligned} Var(\theta) &= E[Var(\theta|x)] + Var[E(\theta|x)] \\ &= Var(\theta|x) + Var(\bar{\theta}) \\ &= (1-\rho)\sigma^2 + \rho\sigma^2 \\ &= \sigma^2. \end{aligned} \tag{14}$$

From Equation 11, the variance of MLEs is an overestimate of the variance of θ. From Equation 13, the variance of Bayesian estimates is an underestimate. In both cases, estimating σ^2 from the variance of a large sample of individual estimates requires adjustments. With MLEs, the adjustment implied by Equation 11 is

$$\sigma^2 \approx \hat{Var}(\hat{\theta}) - \sigma_e^2 \tag{15}$$

if an estimate of σ_e^2 is available, or, equivalently,

$$\sigma^2 \approx \rho\,\hat{Var}(\hat{\theta}) \tag{16}$$

if an estimate of ρ is used. With Bayes mean estimates, Equation 13 implies

$$\sigma^2 \approx Var(\bar{\theta})/\rho. \tag{17}$$

Question 2: Because calculating Bayesian estimates typically requires knowing the population distribution, how should one proceed if it is not known?

Bayesian estimates under the normal-distribution CTT case require the structural parameters μ, σ^2, and σ_e^2. If these are not known, they can be approximated in familiar ways: Equation 10 for an estimate of μ, an internal consistency estimate for ρ, then Equation 16 followed by Equation 8 for estimates of σ^2 and σ_e^2. This section derives an alternative approach that lends itself better to an IRT analogue. The basic idea is first to construct Bayesian estimates for θs by using provisional values for μ and σ^2, and then to employ the mean and variance of the resulting estimates to obtain improved values for μ and σ^2. These values can be used in turn to construct improved estimates for individual examinees.

The provisional values for μ and σ^2 may be denoted by μ^* and σ^{*2}. *Assuming σ_e^2 to be known*, one defines the following quantities:

$$\rho^* = \frac{\sigma^{*2}}{\sigma^{*2} + \sigma_e^2}$$

and

$$\bar{\theta}^* = \rho^* x + (1 - \rho^*)\mu^*.$$

The expected mean and variance of $\bar{\theta}^*$ in the population of examinees, denoted subsequently as M and S^2, are derived as follows:

$$\begin{aligned} M &\equiv \mathrm{E}(\bar{\theta}^*) \\ &= \mathrm{E}[\rho^* x + (1 - \rho^*)\mu^*] \\ &= \rho^* \mathrm{E}(x) + (1 - \rho^*)\mu^* \\ &= \rho^* \mu + (1 - \rho^*)\mu^* \end{aligned} \tag{18}$$

and

$$\begin{aligned} S^2 &\equiv \mathrm{Var}(\bar{\theta}^*) \\ &= \mathrm{Var}[\rho^* x + (1-\rho^*)\mu^*] \\ &= \rho^{*2}\,\mathrm{Var}(x) \\ &= \rho^{*2}\left(\sigma_e^2 + \sigma^2\right). \end{aligned} \tag{19}$$

Given (estimates of) M and S^2, one can then solve for μ and σ^2 in terms of known quantities:

$$\mu = \left[M - (1-\rho^*)\mu^*\right]/\rho^* \tag{20}$$

and

$$\sigma^2 = S^2/\rho^{*2} - \sigma_e^2. \tag{21}$$

These relationships require the existence of the moments that are involved, but not normality.

Question 3: What if Bayesian ability estimates have been calculated in accordance with a common population distribution, but it is later desired to estimate the distributions of specified subpopulations?

Bayesian ability estimation can combine examinees' observed scores with information from other sources, such as a subpopulation membership. Suppose, for example, that the distributions of girls and boys are $\mathrm{N}(\mu_g, \sigma_w^2)$ and $\mathrm{N}(\mu_b, \sigma_w^2)$ respectively—normal, with a common within-group variance. If $\mu_g > \mu_b$, then the Bayes estimate for a girl with a given observed score will be higher than that of a boy with the same score. This might be the way to bet, but it is *not* the way to run a fair contest, such as awarding benefits to individuals. If Bayesian estimates are to be used at all in such a situation, they should be calculated with the same prior distribution for all examinees, so as to preserve rank orderings. But if individual Bayesian estimates based on a common prior are calculated for such purposes, it follows from the preceding section that they will yield biased estimates of subpopulation characteristics when analyzed as if they were true θs. Specifically, the overall population mean and variance play the role of μ^* and σ^{*2} in the preceeding section; the actual mean and variance of a subpopulation of interest correspond to μ and σ^2; and the resulting biased estimates correspond to M and S^2.

As an illustration, the running example of girls and boys is continued. It is assumed that both subpopulations are of equal size, and that Δ denotes the mean difference $\mu_g - \mu_b$. The overall population mean and variance are

$$\mu = (\mu_g + \mu_b)/2$$

and

$$\sigma^2 = \Delta^2/4 + \sigma_w^2.$$

Although the population is actually the mixture of two normals rather than normal itself, Equations 7 through 9 might be employed to approximate the posterior mean and variance for each individual boy and girl. The mean of these Bayes mean estimates for girls is obtained via Equation 18 as a weighted average of the correct value, μ_g, and the overall mean, μ :

$$M_g = E(\bar{\theta}^* | \text{girl}) = \rho\mu_g + (1-\rho)\mu. \tag{22}$$

Equation 22 shows that the degree of bias depends on ρ. An improved estimate of the true girls' mean can be based on Equation 20:

$$\mu_g = [M_g - (1-\rho)\mu]/\rho.$$

Item Response Theory

The essential ideas of IRT are that the probabilities of multiple responses from an examinee are driven by an unobservable proficiency variable θ, and that responses are independent given θ. The 2-parameter logistic IRT model for binary (correct/incorrect) test items, for example, gives the probability of a correct response to Item j as the following function of θ:

$$P(x_j = 1 | \theta, a_j, b_j) = \Psi[a_j(\theta - b_j)], \tag{23}$$

where Ψ denotes the logistic distribution function, $\Psi(z) = [1 + \exp(z)]^{-1}$; a value of 1 for x_j means "correct" and 0 means "incorrect;" and a_j and b_j are parameters of Item j, indicating its sensitivity and difficulty. It is assumed in this presentation that item parameters are known. In practice, of course, they must be estimated. The interested reader is referred to Tsutakawa and Johnson (1990) for one technique for taking uncertainty about item

parameters into account when estimating θ. If the item parameters are estimated accurately, however, this source of uncertainty can be ignored.

Under the usual IRT assumption of conditional, or local, independence, the probability of a vector of responses $\boldsymbol{x} = (x_1, \ldots, x_n)$ to n items is a product of terms over items:

$$P(\boldsymbol{x}|\theta) = \prod_{j=1}^{n} P_j(\theta)^{x_j} Q_j(\theta)^{1-x_j}, \tag{24}$$

where $P_j(\theta) \equiv P(x_j = 1|\theta)$ and $Q_j(\theta) \equiv 1 - P_j(\theta) = P(x_j = 0|\theta)$.

Ability Estimates for Individual Examinees

After $\boldsymbol{x}$ has been observed, Equation 24 is interpreted as a likelihood function $\ell(\theta|\boldsymbol{x})$, and serves as a basis for estimating θ. The maximizing value, again denoted $\hat{\theta}$, is the MLE. For samples of $\boldsymbol{x}$ with fixed θ and large n, $\hat{\theta}$ is approximately normally distributed:

$$\hat{\theta} \sim N(\theta, \sigma_e^2), \tag{25}$$

where the estimation error variance is approximated by the reciprocal of the information function, I_θ:

$$I_\theta = \sum_j \frac{P_j''(\theta)}{P_j(\theta)Q_j(\theta)}, \tag{26}$$

with $P_j''(\theta)$ denoting the second derivative of $P_j(\theta)$ with respect to θ. It should be noted that in contrast to the CTT setting, the sampling variance of the MLE depends on the value of θ. In practice, estimated standard errors are often obtained by evaluating Equation 26 with the $\hat{\theta}$ that corresponds to an examinee's $\boldsymbol{x}$. Their squares, estimated error variances, may be denoted by σ_e^2. Large-sample properties offer no guarantee of distributional properties of $\hat{\theta}$ when n is small, however, and even 80 items can be "small" in unfavorable circumstances:

- The likelihood functions under the one-, two-, and three-parameter logistic IRT models have no finite maxima if all the responses are correct or all are incorrect.

- The likelihood functions under the three-parameter models have no finite maxima for many response patterns with few correct responses, in comparison with the sum of the lower-asymptote item parameters.

- Even when finite maxima exist under the three-parameter model, likelihood functions can be decidely non-normal—often skewed right, sometimes multimodal (Yen, Burkett, and Sykes, 1991).

- Even when likelihoods are roughly normal, the value provided by Equation 26 may not be a good approximation of the inverse of the sampling variance of θ.

As in CTT, Bayesian IRT estimates of θ are obtained via Bayes theorem as measures of center tendency in the posterior distribution, namely $p(\theta|x) \propto \ell(\theta|x)\, p(\theta)$. Bock and Mislevy (1982) outlined numerical approximations for Bayes mean estimation in the context of IRT. One calculates the values of $\ell(\theta|x)$ and $p(\theta)$ at each point along a grid, takes the products at each point, and rescales the results to sum to one. This procedure yields a discrete approximation of the (possibly quite non-normal) posterior $p(\theta|x)$. Its mean and variance are obtained by formulas for weighted means and variances, with the points in the grid serving as observations and their respective posterior probabilities as weights. The resulting Bayes mean estimates and posterior variances, $\bar{\theta}$ and $\bar{\sigma}^2_x$, can be approximated as accurately as desired by spacing the grid points closely enough, and the circumstances described previously that plague maximum likelihood estimation present no such problems. The relevant formulas are shown below, with the grid points denoted Θ_m, for m=1,...,M:

$$p(\Theta_m|x) = \ell(x|\Theta_m)p(\Theta_m) \Big/ \sum_s \ell(x|\Theta_s)p(\Theta_s), \tag{27}$$

$$\bar{\theta} = \sum_m \Theta_m p(\Theta_m|x), \tag{28}$$

and

$$\bar{\sigma}^2_x = \sum_m \left(\Theta_m - \bar{\theta}\right)^2 p(\Theta_m|\xi). \tag{29}$$

If $p(\theta)$ were normally distributed, as in Equation 1, and if an asymptotic normal approximation could be obtained for $\hat{\theta}$ via Equation 25, an examinee's Bayesian mean

estimate and posterior variance could be approximated by revising Equations 7 through 9 as follows:

$$\theta|x,\mu,\sigma^2 \sim N\left(\bar{\theta},\bar{\sigma}_x^2\right), \tag{30}$$

with

$$\begin{aligned}\bar{\sigma}_x^2 &= \left(\sigma^{-2}+\sigma_x^{-2}\right)^{-1} \\ &= (1-\rho_x)\sigma^2,\end{aligned} \tag{31}$$

$$\rho_x = \frac{\sigma^2}{\sigma^2+\sigma_x^2}, \tag{32}$$

and

$$\bar{\theta} = \rho_x\hat{\theta} + (1-\rho_x)\mu. \tag{33}$$

The preceeding formulas apply as approximations for those examinees with response patterns yielding finite values for $\hat{\theta}$ and σ_x^2. Were this the case for *all* response patterns in a data set, one could calculate the average error variance, and then apply the formulas in the CTT sections to approximate population and subpopulation parameters. For examinees infinite MLEs, however, Equations 30 through 33 cannot be applied. Because Bayesian estimates can be obtained for all patterns, however, it may be useful to use them as the basis for approximating for the population mean and variance. To motivate the approximations, *direct* maximum likelihood estimation of population parameters—that is, bypassing the step of estimating individuals' θs —is first reviewed.

Estimates for Population Parameters

The expression $X=(x_1,\ldots,x_N)$ may be used to denote the response vectors from a sample of N examinees. If $\theta \sim p(\theta|\alpha)$, where α is the possibly vector-valued parameter of the distribution, the maximum likelihood estimate of α is obtained by maximizing the marginal likelihood function

$$\ell(X|\alpha) = \prod_i \int p(x_i|\theta)p(\theta|\alpha)\partial\theta. \tag{34}$$

One obtains the maximum by setting to zero the first derivatives of the natural logarithm of Equation 34 with respect to each of the elements of $\boldsymbol{\alpha}$, and then finding the values that solve these resulting likelihood equations (Mislevy, 1984). If $p(\theta|\boldsymbol{\alpha})$ is the univariate normal density, for example, then $\boldsymbol{\alpha}=(\mu,\sigma^2)$. Whether or not normality is assumed for the θ distribution, the maximum likelihood estimates of the population mean and variance can be written in terms of the posterior means and variances of the individual examinees:

$$\begin{aligned}\hat{\mu} &= \sum_i \mathrm{E}(\theta|\boldsymbol{x}_i,\hat{\boldsymbol{\alpha}}) \\ &= \sum_i \bar{\theta}_i\end{aligned} \tag{35}$$

and

$$\begin{aligned}\hat{\sigma}^2 &= N^{-1}\sum_i \mathrm{Var}(\theta|\boldsymbol{x}_i,\hat{\boldsymbol{\alpha}}) + N^{-1}\sum_i \left[\mathrm{E}(\theta|\boldsymbol{x}_i,\hat{\boldsymbol{\alpha}}) - \hat{\mu}\right]^2 \\ &= N^{-1}\sum_i \bar{\sigma}_i^2 + N^{-1}\sum_i \left[\bar{\theta}_i - \hat{\mu}\right]^2.\end{aligned} \tag{36}$$

That is, the MLE of μ is the mean of the Bayesian estimates of the examinees, and the MLE of σ^2 is the sum of the posterior variances and the variance of the posterior means—*provided that they were calculated with the correct mean and variance at the start.* The results specialize to Equations 12 and 14 in the case of CTT. Mislevy (1984) shows how this property of "self-consistency" lies at the heart of estimating $\boldsymbol{\alpha}$ by means of Dempster, Laird, and Rubin's (1977) EM algorithm .

Approximations Based on Bayesian Estimates

One can begin with an provisional approximation for $p(\theta)$, which may, but need not be, normal. Initial values for the mean and variance may be denoted by μ^* and σ^{*2}. An improved approximation of μ and σ^2 is obtained by modifying the CTT correction formulas as follows:

1. Obtain Bayes mean estimates and posterior variances, $\bar{\theta}_i$ and $\bar{\sigma}_i^2$, i=1,...,N, for all examinees.

2. Calculate M and S^2, the sample mean and variance of the $\bar{\theta}_i$ s.

3. Calculate the average of the individual examinees' posterior variances:

$$\overline{\sigma}^2 = N^{-1} \sum_i \overline{\sigma}_i^2. \tag{37}$$

4. Calculate a psuedo-average error variance, analogous to σ_e^2 in the CTT solution:

$$\sigma_e^{*2} = \left(\overline{\sigma}^{-2} + \sigma^{*-2} \right)^{-1}. \tag{38}$$

5. Calculate a psuedo-average reliability coefficient:

$$\rho^* = \frac{\sigma^{*2}}{\sigma^{*2} + \sigma_e^{*2}}.$$

6. Apply Equations 20 and 21 to obtain improved approximations of μ and σ^2. Analogous formulas can be used to approximate subpopulation means and variances when a common mean and variance was used to generate the original set of estimates for individuals.

A Numerical Illustration

This example is based on the responses of 325 students to a 19-item test. The items were open-ended, and the two-parameter logistic model was fit to the data with Mislevy and Bock's (1983) BILOG program. The scale was set so that the mean and variance of the sample were 0 and 1 respectively. The approximation formulas of the preceeding section were employed, starting from values for μ^* of -1, 0, and 1, crossed with values for σ^{*2} of .25, 1, and 4. From the resulting improved estimates in each combination, a second approximating step was then carried out. The results are shown in Figure 1.

[Figure 1 about here]

Each panel in Figure 1 contains the following values:

- Provisional estimates at the start of an approximation cycle, μ^* and σ^{*2}. With these, Bayesian posterior means and variances were calculated for all examinees using BILOG.

- Intermediate calculations M, S^2, ρ^*, and σ_e^{*2}, which are functions of μ^* and σ_e^{*2} and the estimates of provisional posterior means and variances for individual examinees based on μ^* and σ_e^{*2}.

- The resulting updated estimates $\hat{\mu}$ and $\hat{\sigma}^2$.

The center panel starts from, and returns to, the MLE values of 0 and 1. The panels around the perimeter correspond to initial values for μ of -1, 0, or 1, and for initial values for σ^{*2} of .25, 1, or 4. The resulting improved estimates were used in turn for a second adjustment cycle, summaries of which appear in the panel closer next to the center.

Although this example is meant to be illustrative rather than comprehensive, some tentative observations can be made from the results. In each case, a single adjustment step produced an accurate estimate of the mean. Even from the initial approximations farthest from the correct value, a single step would have been sufficient. The adjustments also improved the estimates of the population variance in each case, but not by as much (although it may be noted that the results are given in terms of variances rather than standard deviations; standard deviations are off by only about 5-percent). Unless initial approximations are fairly accurate, it would appear prudent to carry out at least two adjust steps in order to obtain a satisfactory approximation of the variance.

References

Bock, R.D., & Mislevy, R.J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement, 6*, 431-444.

Box, G.E.P., & Taio, G.C. (1973). *Bayesian inference in statistical analysis*. Reading, Mass: Addison-Wesley.

Dempster, A.P, Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B, 39*, 1-38.

Kelley, T.L. (1923). *Statistical methods*. New York: Macmillan.

Lord, F.M. (1969). Estimating true score distributions in psychological testing (An empirical Bayes problem). *Psychometrika, 34*, 259-299.

Mislevy, R.J. (1984). Estimating latent distributions. *Psychometrika, 49*, 359-381.

Mislevy, R.J., Beaton, A.E., Kaplan, B., & Sheehan, K.M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement, 29*, 133-161.

Mislevy, R.J., & Bock, R.D. (1983). *BILOG: Item analysis and test scoring with binary logistic models* [computer program]. Mooresville, IN: Scientific Software, Inc.

Novick, M.R., & Jackson, P.H. (1974). *Statistical methods for educational and psychological research*. New York: McGraw-Hill.

Tsutakawa, R.K., & Johnson, J. (1990). The effect of uncertainty of item parameter estimation on ability estimates. *Psychometrika, 55*, 371-390.

Yen, W.M., Burket, G.R., & Sykes, R.C. (1991). Nonunique solutions to the likelihood function for the three-parameter logistic model . *Psychometrika, 56*, 39-54.

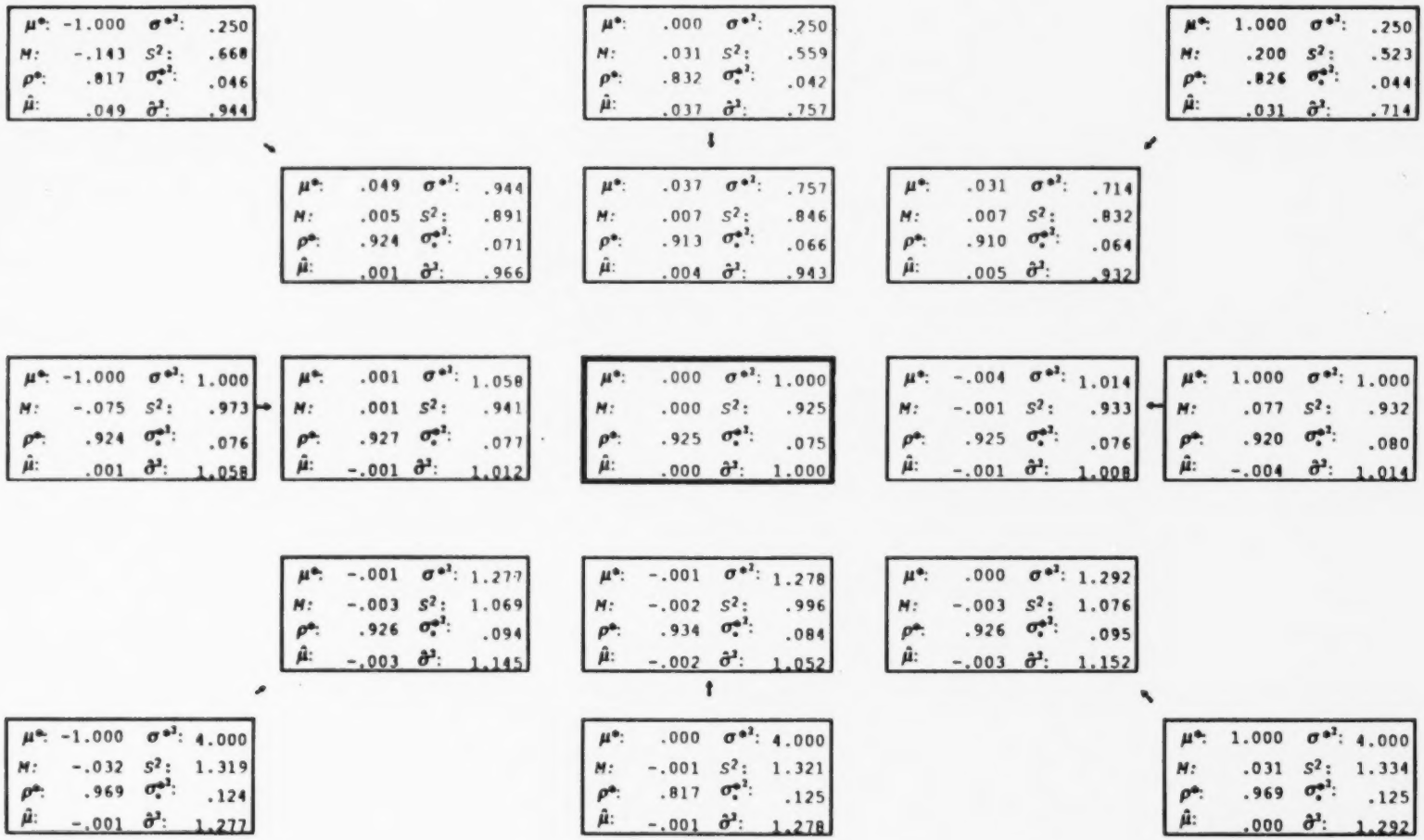

Figure 1

Illustrative Results of Adjustment Formulas in the Context of IRT